CMPU-250 Project Group 1
Peer Review for Group 3
4/18/2025

Introduction

The introduction did well in providing a solid background context for the paper. Providing the context of how Food Price Estimate information was gathered in different countries, leading to different visibilities was really important in understanding the paper. This context provided a strong motivation for the why of creating and analyzing both a new model and the current World Bank Food Price Estimate, the research questions are quite interesting and definitely something worth exploring. The portion of text taken from the World Bank was a good synopsis and quite helpful to general readers in understanding the purpose of the organization that the paper will then leverage to talk about the algorithms they use to get these estimates. The citations and related literature were a good touch and very thorough. Overall, the introduction provided a solid background, and strong research questions to be further explored further in the paper.

Data

The source of the data was properly provided coming from the World Bank's Microdata Library, and further specified by which dataset was used "Monthly food price estimates by product and market." The time frame of when (2007-2025) and how the data was collected was also stated. They also stated the variables that were removed for being irrelevant to the subject of the paper; such as geospatial coordinates, country codes, and extraneous date data. The data appears to be clean and tidy. Furthermore a figure is provided with a specific case with bread in Afghanistan to show how the data is being gathered. Overall, the data is well contextualized and cleaned. It might be nice to provide an explanation of what a tier 1/ tier 2 administrative district is, for the general reader to better understand the meaning of that in the data section.

Methods

Overall, the methods section was well-done and well explained. They were very clear where they got their data from, as well as how they performed the data cleanup and model training. Any additional variables they added to the dataset were very well explained. Their methods fit their questions well, and help answer their research questions. The three countries they chose to analyze (Afghanistan, South Sudan, and Somalia) all were very well detailed, and the accompanying graphs were labeled and easy to read. We thought the heatmaps were particularly insightful, as they helped give an overview into what goods suffered price fluctuations. Given that they also wanted to see what external factors (climate crises, natural disasters, conflicts, etc.)affected commodity prices, we think that they should include how they intend to account for these variables in their model. We would also like to see how they would like to expand their cases beyond these three countries, and how they would tailor or change their methodology.

Results

Their results are well explained, and each section clearly explains their findings in their three case countries. Based on the results, their model fits the question well, and helps account for

their findings. They clearly state trends and timelines, which are important for their model. We do think they should try to explain their findings more, either here or in a separate discussion section. We think the paragraphs on South Sudan and Somalia could have some elaboration, particularly in regard to external factors like the ones stated in the methods section.

Discussion

There is no discussion section yet but we talked with the group about possible directions for relating their results to their research questions. With their brief discussion of the war in Afghanistan in the results section (which can eventually be moved to the discussion section), they begin to look into the possible shocks that might disrupt price predictions. We suggest they also deepen this discussion and discuss similar shocks in South Sudan and Somalia. We also suggest looking into how the World Bank price predictions are used by NGOs and governments (likely for aid distribution) in order to assess the possible impacts of misprediction (allocative harms under Shelby et al.'s taxonomy).

Analysis Code

The code used for analysis looks correct. The code is relatively readable and commented, but could use some style consistency. Some strings are double-quoted while others are single-quoted. There are also some missing spaces around some operators. Additionally, there is a lot of repeated code between the data cleaning, preparation, and analysis for the three countries which seems like it could be abstracted into a common function or common functions.

There are no markdown descriptions of the analysis, but some of the descriptions from the PDF could be easily copied to provide analysis of the graphs. It would also be interesting to see what metrics you used (even if it was just looking at the density of the data points) to determine which commodities to use reflected in the notebook. The output of all the code blocks are accessible.

General

The writing is generally clear, and effectively conveys the work done on the project. In some places, the grammar and phrasing could be changed to eliminate repetition and improve the flow. (On page 6, for example, the introduction to a visualization repeats that the visualization represents the average percent difference in the price of beans in South Sudan, and in a few other places, there are missing commas. Nothing major; just worth a quick proofread when you revise the report.) The tone could be made slightly more formal in some places, such as when (also on page 6) the report reads "We did a similar thing to figure out if there is a difference…" The code is also clear, and it is evident how you truncated the data, analyzed it, and created your visualizations. Everything appears reproducible. The report is well-formatted, and sticks to relevant analysis and output; the visualization which is a screenshot of DataViz may not be necessary, but maybe it's important to illustrate the data collection process. The next steps seem to be adding a discussion section and delving deeper into the ramifications of your analysis. These next steps could have been more explicitly discussed in the report, but generally it is clear that you have a vision for where the project is going. In terms of organization, the repository appears to fit the requirements laid out, and it was easy to find files.

The vision for the project moving forward seems clear, but we had a few suggestions for future direction and changes that might make your analysis even stronger. Because your project is so tied to geography, examining food prices in different locations in different countries, some visualizations reflecting this would be a nice touch. More explicit discussion of geographic differences would also be helpful, as readers may not understand the distinction between different South Sudanese markets. Other helpful details could include adding pointers on timescale visualizations to make it clear when major events occurred and how they may have impacted the prices and predictions, and perhaps looking at the prices of different commodities within each country to see if they were all affected equally.

Final Considerations

One question we had after reading your analysis was: what do you see as the potential ramifications of your project? What conclusions would you say are important for those working on the model you're assessing, or any other food price prediction model, to understand?

In terms of what was done well—the report was generally effective and well done, but we were particularly impressed with the heatmap visualizations. They managed not only to convey information about the changing commodity prices and attendant trends, but also about the nature of the data itself, clearly showing where data was and was not available about prices. In this way, we found them doubly effective in elucidating both the analysis itself and the data on which it was based.